A Space and Solar Physics Great Observatory: Virtual but Nearing Reality

Aaron Roberts
NASA Goddard Space Flight Center
Laboratory for Solar and Space Physics, Code 612.2
Greenbelt, MD 20771

Vasili Rezapkin
Aquilent
1100 West St.
Laurel, MD 20707

Joseph King
QSS, Inc.
4500 Forbes Blvd.
Lanham, MD 20706

Abstract **Virtual Observatories represent a new paradigm for research in many fields. Recent advances in this area for Space and Solar Physics have led to operational systems and the promise of a highly integrated research environment in which a fleet of spacecraft and ground facilities function as a virtual "great observatory."**

I. WHY A GREAT OBSERVATORY?

Progress in space and solar physics increasingly depends on multispacecraft and spacecraft/model integration. We now have many resources online, but in many places and in different formats. To solve problems such as those posed by NASA's Living With a Star Program, a new paradigm is needed in which the various observational and computational resources are easily accessible in a uniform way. This is the idea behind a (Virtual) Sun-Solar System Great Observatory (VS3O).

II. DEFINITION OF A "VIRTUAL OBSERVATORY"

The original definition of a Virtual Observatory is as an electronic version of ground-based astronomical observatories that allows researchers to examine any sky region in multiple wavelengths from a desktop computer using stored data at many sites. By analogy with this effort, and generalizing it for our goals, we define a Virtual Observatory (VO) as a suite of software applications on a set of computers that allows users to uniformly find, access, and use resources (data, software, document, and image products and services using these) from a collection of distributed product repositories and service providers. A VO is a service that unites multiple services and/or repositories.

III. PURPOSES OF VOS

There are two main purposes for Virtual Observatories. First, they should make "standard" scientific research much more efficient. Ideally, even the PI teams should want to use them to get to their own data because of the ease-of-use and enhanced services. To accomplish this goal, VOs must improve on existing services such as mission and PI sites and conventional data centers. VOs will not replace these, but will use them in new ways.

The second purpose of VOs is to enable new, global problems to be solved by providing discovery, access, and analysis tools that allow users to rapidly gain integrated views from the solar origin of a phenomenon to its terrestrial effects. Scientists need to find data related to any particular observation and ultimately answer "higher-order" queries such as "Show me the data, in a 3-D view, from cases where a large Coronal Mass Ejection observed by the SOHO spacecraft was also observed in situ near the Earth."

IV. THE STRUCTURE OF A VO DATA ENVIRONMENT

The core of a VO data environment is a large set of high-quality data and model products that are readily available along with the software to use them. Automated data reduction is key to making such products. Open, public availability makes the user part of the data-quality effort; for some organizations this is a new paradigm. An open-data policy is generally good for the providers.

For easy access and use, the resources (products and services) need to be electronically registered in a uniform language. A community accepted "data model" provides the required semantics. (See the SPASE data model at http://www.igpp.ucla.edu/space.) A general registry with web-based entry and easy export should keep track of additions and changes. Product IDs will help a great deal to simplify access and use. Providers must offer direct machine access, preferably using the above uniform language, to the resources. Simple APIs make access easier. A common set of protocols (e.g., "SOAP", cgi, and/or ftp standards) will let users and other services access a wide variety of resources. "VxOs" can organize subfields. "Gateways" or "brokers" running as applications on server machines enable finding and accessing the resources using the registry and the access methods. The gateways, in turn, should have APIs that allow applications to treat them as an integrated resource provider. Based on the API, a variety of front-end applications, including simple browser applications and scientific analysis programs, will allow access to the resources registered in the gateway. The architecture is flexible, and enables many variations on the above scenario.

At a higher level, the basic VO just described makes possible the development of new applications and services to use data products. Applications can be downloaded, repository provided (e.g., graphics and subsetting), or web service accessible (e.g., "SolarSoft" via "CoSEC"). Some useful services are: running "on-demand" models; reading, displaying, and translating multiple formats; transforming coordinates; merging related datasets and performing correlations; and browsing data plots. A commonly desired service is one that will allow "higher-order" queries using "pre-mining" of data to produce event lists and modest resolution datasets. We want to ask things like, "When were there solar flares, strong geomagnetic activity, and spacecraft both upstream of the Earth and in the magnetotail?" Direct data mining can also be automated using integrated or grid-based parallel computation to handle large data volumes.

V. The Virtual Space Physics Observatory

There are a few VOs in Space and Solar Physics that are now operational, while others are planned. A number of "VxOs" intending to provide access to products from subfield "x," are in development. The first and farthest along of these is the Virtual Solar Observatory, which provides direct access to a wide variety of solar data products from many observatories, mostly in the form of files in the widely used self-documenting FITS format. Other similar efforts are underway for Heliospheric, ITM, Geomagnetic, and Radiation Belt Observatories. A great deal of information on the background on these projects and links to these and other sites can be found at http://lwsde.gsfc.nasa.gov. This site also includes a "White Paper" that gives a more detailed description of VOs and the related data environment: http://lwsde.gsfc.nasa.gov/VO_Framework_7_Jan_05.pdf (a MS Word version is also available on the site).

The VSO provides very uniform access to data in a broad subdiscipline. In contrast, our Virtual Space Physics Observatory (VSPO) focuses on providing access to a wide variety of data, image, model, and other products, although less uniformly and often somewhat less directly than VSO. We provide access to products or URLs to sites that deliver these products as directly as possible. A dynamic web page front-end allows "Google-like" and element-based searches for products. "SOAP" interfaces pass internal messages, and are used for direct repository access. Recently we have added access to solar products using the VSO API to provide the products they deliver in a complementary way. Our access currently has more metadata about the products, and can be more direct, but we have not implemented all of the VSO features such as being able to retrieve multiple types of products at the same time. VSPO includes additional service links for specific orbit information, journal references, and current space weather for convenience.

The internal structure and Web-based interface to VSPO are shown in the figures below.

VI. Future Directions

We intend to extend the VSPO to access more products by more means, and to link it to services that will allow many data products to be treated uniformly by a set of plotting and analysis tools. We will also produce and link VSPO to a companion higher-order query service based on event lists and modest resolution data sets held and manipulated in RAM. This will be done in collaboration with a large number of people and groups who will be developing repository access methods and other ways of accessing and using the data, in addition to the SPASE group that is defining a Solar and Space Physics Data Model. We hope that this integrated approach will lead to a genuinely useful Great Observatory that will provide the basis for a new level of global-scale scientific research.

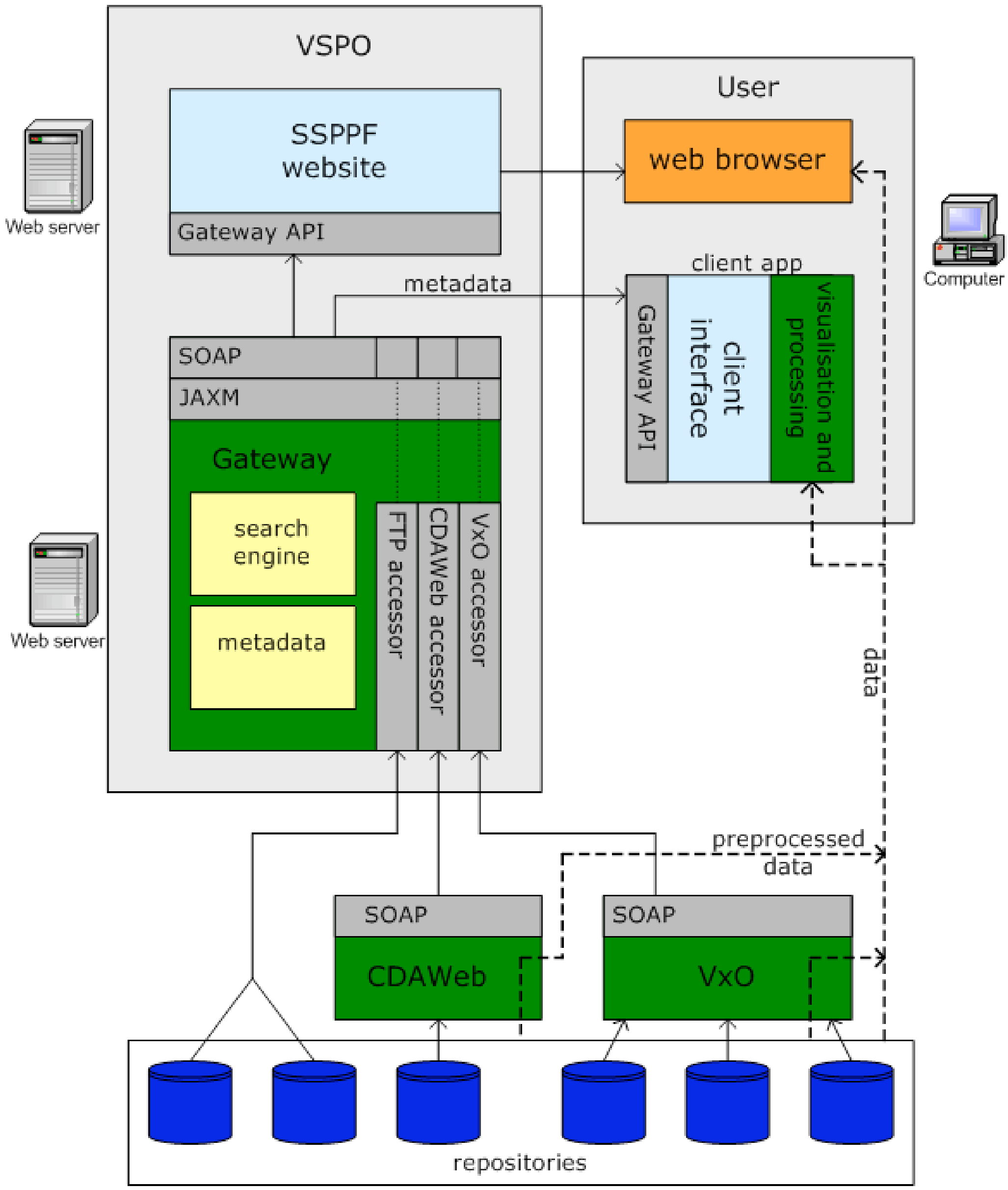

Fig. 1: The basic architecture of VSPO. A Gateway stores a metadata registry of products that are requested through "accessors" and delivered to the user outside of VSPO. The general framework presented here is typical of most VOs in its three-part front-end, broker, repository structure, although there are many variations in the details.

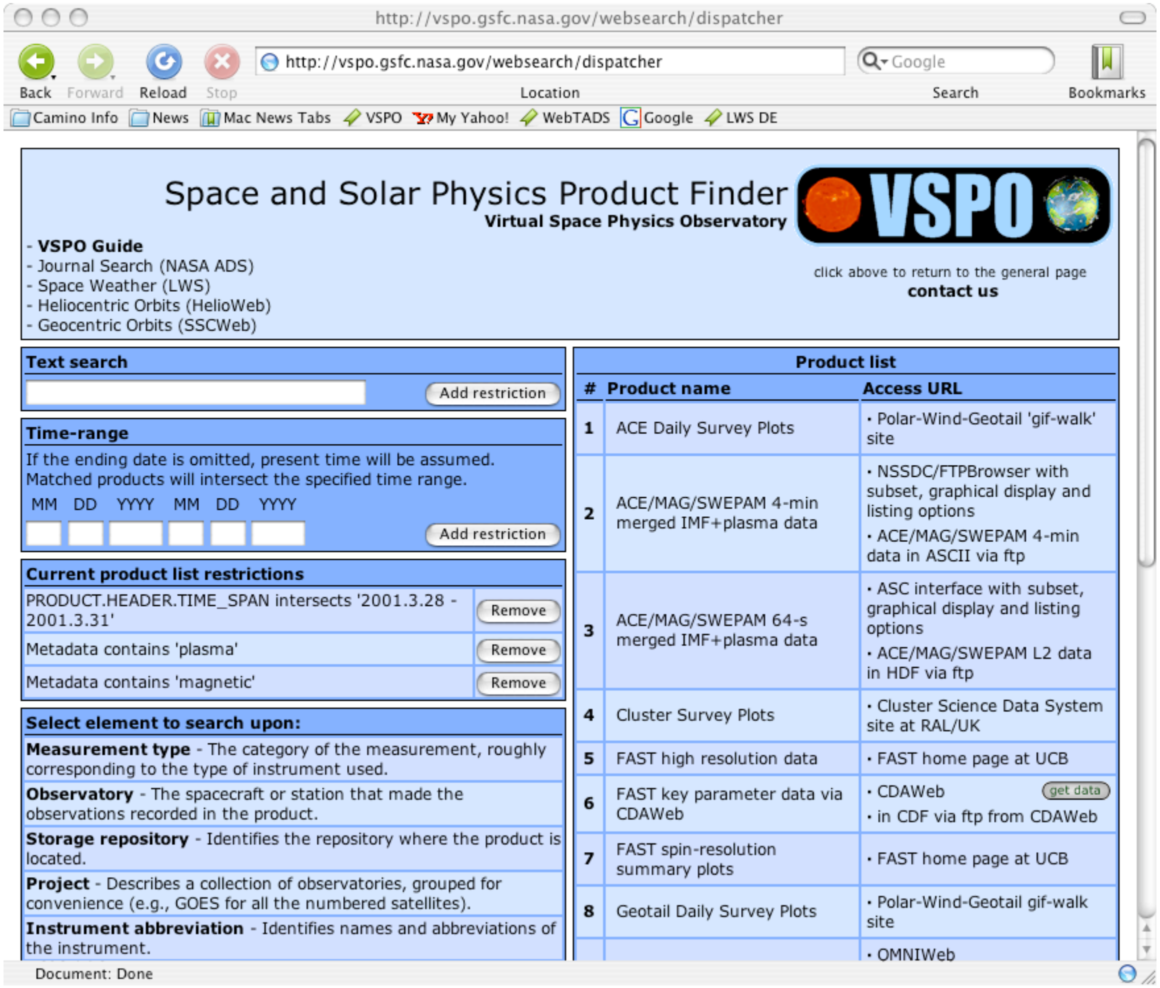

Fig. 2: The current Web browser access to VSPO (see http://vspo.gsfc.nasa.gov). Note the "Google-like" search box, the ability to search based on many data model terms, and the list of currently selected products and links to URLs and data accessors.